

What Is Claimed Is:

1. A speech recognition system comprising:

5 microphone means for receiving acoustic waves and converting the acoustic waves into electronic signals;

linear prediction (LP) signal processing means, coupled to said microphone means, for processing the electronic signals to generate LP parametric representations of the electronic
10 signals;

mel-frequency linear prediction (MFLP) generating means, coupled to said LP signal processing means, for mel-frequency warping said LP parametric representations to generate MFLP parametric representations of the electronic signals; and
15

word comparison means coupled to said MFLP means, for comparing said MFLP parametric representations of the electronic signals to parametric representation of words in a database.

20 2. The speech recognition system of claim 1, wherein said mel-frequency linear prediction (MFLP) generating means comprises:

non-uniform discrete Fourier transform (NDFT) generator means for generating the NDFT of said LP parametric representations of the electronic signals;
25

warper means, coupled to said NDFT generator means, for mel-frequency warping said NDFT;

smoothing means, coupled to said warper means, for smoothing said mel-frequency warped NDFT; and
30

cepstral parameter converter means, coupled to said smoothing means, for converting said LP parametric representations of the electronic signals to cepstral parameters.

5 3. The speech recognition system of claim 2 wherein said smoothing means utilizes a low-order all-pole LP generator.

4. The speech recognition system of claim 1 wherein said word comparison means is a dynamic time warper speech recognition system.

10

5. The speech recognition system of claim 1 wherein said word comparison means is a hidden Markov model speech recognition system.

6. The speech recognition system of claim 1 wherein said word comparison means is a neural network speech recognition system.

15

7. A speech recognition system for recognizing a speech signal, comprising:

a pre-emphasizer for spectrally flattening the speech signal;

a frame blocker, coupled to said pre-emphasizer, for frame blocking the speech signal;

a windower, coupled to said frame blocker, for windowing each blocked frame;

20

25 a pre-warp LP generator, coupled to said windower, to generating a plurality of pre-warp LP parameters;

a mel-NDFT warper, coupled to said pre-warp LP generator, for utilizing a non-uniform discrete Fourier transform (NDFT) to warp said pre-warp LP parameters on a mel scale to generate a plurality of mel scale-warped LP parameters;

30

a power spectrum generator, coupled to said mel-NDFT warper, for generating a warped vocal-tract power spectrum from said mel scale-warped LP parameters;

5 an IDFT generator, coupled to said power spectrum generator, for generating an inverse discrete Fourier transform of the warped vocal-tract power spectrum;

a post-warp LP generator, coupled to said IDFT generator, for generating a plurality of post-warp LP parameters; and

10

a cepstrum converter, coupled to said post-warp LP generator, for converting said post-warp LP parameters to a plurality of MFLP cepstral coefficients.

15
20
25
30
35
40
45
50
55
60
65
70
75
80
85
90
95
100
105
110
115
120
125
130
135
140
145
150
155
160
165
170
175
180
185
190
195
200
205
210
215
220
225
230
235
240
245
250
255
260
265
270
275
280
285
290
295
300
305
310
315
320
325
330
335
340
345
350
355
360
365
370
375
380
385
390
395
400
405
410
415
420
425
430
435
440
445
450
455
460
465
470
475
480
485
490
495
500
505
510
515
520
525
530
535
540
545
550
555
560
565
570
575
580
585
590
595
600
605
610
615
620
625
630
635
640
645
650
655
660
665
670
675
680
685
690
695
700
705
710
715
720
725
730
735
740
745
750
755
760
765
770
775
780
785
790
795
800
805
810
815
820
825
830
835
840
845
850
855
860
865
870
875
880
885
890
895
900
905
910
915
920
925
930
935
940
945
950
955
960
965
970
975
980
985
990
995

8. The speech recognition system of claim 7 wherein said pre-emphasizer is a fixed low-order digital filter.

9. The speech recognition system of claim 7 wherein said windower is a Hamming window.

10. The speech recognition system of claim 7 wherein said warped vocal-tract power spectrum is modeled utilizing a predetermined number of peaks.

11. The speech recognition system of claim 7 further comprising:

25 a word template for storing a plurality of cepstral coefficient parametric representations of word pronunciations;

a dynamic time warper for dynamic behavior analysis of said MFLP cepstral coefficients; and

30

a word comparator, coupled to said cepstrum converter, to said word template, and to said dynamic time warper, for comparing said plurality of MFLP cepstral coefficients with said plurality of cepstral coefficient parametric representations of word pronunciations;

5 12. A mobile communication device comprising:

a flash memory;

a microprocessor, coupled to said flash memory,

10

a DSP processor, coupled to said flash memory and said microprocessor, and responsive to said flash memory and said microprocessor, for performing mel-frequency linear prediction (MFLP) speech recognition;

15

a read-only-memory (ROM) device, coupled to said DSP processor, for storage of data; and

a random access memory (RAM) device 505, for storage of data.

20

13. A method for modifying the linear prediction (LP) vocal-tract spectrum comprising the steps of:

(a) mel-frequency warping the LP vocal-tract spectrum to generate a mel-frequency warped LP vocal-tract spectrum;

25 (b) modeling said mel-frequency warped LP vocal-tract spectrum utilizing a predetermined number of peaks; and

(c) performing linear prediction on said modeled mel-frequency warped LP vocal-tract spectrum to generate an LP mel-frequency warped LP vocal-tract spectrum.

14. The method of claim 13 wherein step (a) comprises the steps of:

(a) calculating the discrete-time Fourier transform (DTFT) of the finite impulse response LP parameters;

5 (b) taking a predetermined number of samples of said DTFT of the finite impulse response LP parameters;

(c) utilizing a non-uniform grid for said DTFT of the LP vocal-tract spectrum to generate a non-uniform discrete Fourier transform (NDFT); and

10 (d) oversampling a mel filterbank to generate a warped grid for said NDFT of the finite impulse response LP parameters.

15 15. The method of claim 13 wherein said non-uniform grid of step (c) is substantially similar to the mel frequency scale.

16 16. The method of claim 14 wherein said oversampling of step (d) is linear from 0 to 1000 Hz and frequency samples in the octaves greater than 1000 Hz are sampled at equal spaces in the log domain.

17 17. The method of claim 13 wherein said predetermined number of peaks in step (b) is two.

18. The method of claim 13 wherein said step (c) comprises the steps of:

computing the inverse discrete Fourier transform (DFT) said modeled mel-frequency warped LP vocal-tract spectrum;

25 generating a predetermined number of samples of an autocorrelation sequence of said modeled mel-frequency warped LP vocal-tract spectrum; and

performing linear prediction to generate a plurality of LP parameters from said modeled mel-frequency warped LP vocal-tract spectrum.

19. A method for processing speech acoustic signals, comprising the steps of:

- (a) receiving the speech acoustic waves utilizing a microphone;
- (b) converting the speech acoustic waves into electronic signals;
- (c) parameterizing the electronic signals utilizing linear prediction (LP);
- (d) mel-frequency warping said linear prediction parametric representations; and
- (e) comparing said mel-frequency warped linear prediction parametric representation with parametric representations of words in a database.

5

579241v1